Testing source confusion and identification capability in Cherenkov telescope array data

Enrique Mestre [●],[1,2]★ Diego F. Torres,[1,3,4]★ Emma de Oña Wilhelmi [●][5] and Josep Martí[2]

[1]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, E-08193 Barcelona, Spain*
[2]*Departamento de Física (EPSJ), Universidad de Jaén, Campus Las Lagunillas s/n, A3, E-23071 Jaén, Spain*
[3]*Institució Catalana de Recerca i Estudis Avançats (ICREA), E-08010 Barcelona, Spain*
[4]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[5]*Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, D-15738 Zeuthen, Germany*

ABSTRACT

The Cherenkov Telescope Array will provide the deepest survey of the Galactic Plane performed at very-high-energy gamma-rays. Consequently, this survey will unavoidably face the challenge of source confusion, i.e. the non-unique attribution of signal to a source due to multiple overlapping sources. Among the known populations of Galactic gamma-ray sources and given their extension and number, pulsar wind nebulae (PWNe, and PWN TeV haloes) will be the most affected. We aim to probe source confusion of TeV PWNe in forthcoming CTA data. For this purpose, we performed and analysed simulations of artificially confused PWNe with CTA. As a basis for our simulations, we applied our study to TeV data collected from the H.E.S.S. Galactic Plane Survey for ten extended and two point-like firmly identified PWNe, probing various configurations of source confusion involving different projected separations, relative orientations, flux levels, and extensions among sources. Source confusion, defined here to appear when the sum of the Gaussian width of two sources is larger than the separation between their centroids, occurred in ∼30 per cent of the simulations. For this sample and 0.5° of average separation between sources, we found that CTA can likely resolve up to 60 per cent of those confused sources above 500 GeV. Finally, we also considered simulations of isolated extended sources to see how well they could be matched to a library of morphological templates. The outcome of the simulations indicates a remarkable capability (more than 95 per cent of the cases studied) to match a simulation with the correct input template in its proper orientation.

Key words: instrumentation: detectors – ISM: supernova remnants.

1 INTRODUCTION

The Cherenkov Telescope Array[1] (CTA; Cherenkov Telescope Array Consortium 2019) is the next major ground-based observatory for very-high-energy (VHE, $E > 100$ GeV) gamma-ray astronomy. CTA is to be located on both hemispheres, a northern location in La Palma (Spain) and a southern one in Paranal (Chile). Each site will host an array of Small-, Medium-, and Large-Size Telescopes (SSTs, MSTs, and LSTs) sensitive to different energy ranges from 20 GeV to more than 300 TeV. SSTs are currently planned to be installed only at the southern site. CTA will improve the sensitivity of current Imaging Atmospheric Cherenkov Telescopes (IACTs) by a factor from five to ten depending on the energy range (Acharyya et al. 2019). The first Large-Sized Telescope (LST-1; Mazin 2021), built on-site (currently finalizing the commissioning phase) and located at the Observatorio del Roque de los Muchachos (La Palma), has been operating since 2018 October.

The CTA main goals are categorized in various Key Science Projects (KSPs), including surveys (e.g. the Galactic Plane Survey,

GPS) targeting various types of sources. The CTA GPS (Dubus et al. 2013; Cherenkov Telescope Array Consortium 2019) will cover the full Galactic plane, especially the inner region (i.e. $|l| < 60°$), using both the northern and southern arrays at point-source sensitivities of a few mCrab.[2] Its main goal is to provide a census of Galactic VHE gamma-ray sources, identify promising targets for follow-up observations, and characterize the Galactic plane diffuse emission properties.

The High Energy Spectroscopic System (H.E.S.S.; Bernlöhr et al. 2003) experiment released the most comprehensive survey of the Galactic plane at VHE gamma-rays up to date (HGPS; H. E. S. S. Collaboration 2018). The HGPS comprises about 2700 h of data (after quality selection) for longitudes from $l = 250°$ to 65° and latitudes $|b| \leq 3°$. The resulting HGPS catalogue contains 78 VHE sources with 31 firmly identified pulsar wind nebulae (PWNe), supernova remnants (SNRs), composite SNRs, or gamma-

★ E-mail: emestre@ujaen.es (EM); dtorres@ice.csic.es (DFT)
[1] https://www.cta-observatory.org

---

[2]The Crab unit is the flux of the source having a spectrum similar to that measured for the Crab nebula and pulsar between 500 GeV and 80 TeV with HEGRA (dN/d$E = 2.83 \times 10^{-11}(E/1\text{TeV})^{-2.62}$ cm^{-2} s^{-1} TeV^{-1}; Aharonian et al. 2004) above a certain energy. For example, 1 mCrab $= 5.07 \times 10^{-13}$ ph cm^{-2} s^{-1} for an energy threshold of 125 GeV.

ray binaries. The future CTA GPS catalogue is expected to achieve the detection at 5σ significance of $\sim$500 VHE sources at energies from 70 GeV to 200 TeV, including more than 200 PWNe (Remy et al. 2021). It is more than six times the objects in the HGPS or the third High Altitude Water Cherenkov catalog (3HAWC, with 65 TeV sources detected, Albert et al. 2020).

The source confusion problem, i.e. the difficulty in discriminating sources that overlap in crowded regions, is a problem that the CTA GPS will need to address. Particularly given the large extension and number of some gamma-ray source classes and the relatively low angular resolution of IACTs at tens of GeV to TeV energies (i.e. $\gtrsim$ 0.05°). Initial studies led to an approximate lower limit to the amount of source confusion of 13–24 per cent at 100 GeV and 9–18 per cent at 1 TeV in the region defined by $|l| < 30^\circ$ and $|b| < 2^\circ$ (Cherenkov Telescope Array Consortium 2019).

Our main goal is to probe, through simulations, the source confusion problem in forthcoming CTA data and the instrument's identification capabilities. We are particularly interested in constraining the source confusion regarding the Galactic population of TeV PWNe, which is expected to be the most numerous population of sources. For this purpose, we tested if future CTA data can be directly compared to a library of empirical (e.g. here based on HGPS data) morphological source templates. Furthermore, we aim to probe whether the cross-match with such a library could provide hints to unravel source confusion and, if so, to what extent. In the future, this work will pave the way to explore the possibility of employing simulated morphological templates instead (from magnetohydrodynamical; MHD, hydrodynamical; HD, or HD + B results, see e.g. Volpi et al. (2008), Kolb et al. (2017), Olmi & Torres (2020)).

2 SIMULATIONS AND METHODS

2.1 The library of source templates

We first built a library of spatial templates in the form of sky maps, which describe the morphology of different PWNe by depicting the intensity distribution in a region containing each source of interest. We obtained the above-mentioned templates from the H.E.S.S. Galactic plane survey (H. E. S. S. Collaboration 2018). The HGPS events maps (see Section 3.1 of the HGPS paper) are built from the reconstructed positions of the primary gamma-ray photons from all events observed by the H.E.S.S. survey covering the region defined by $l = 70^\circ$ to 250° and $b = \pm 5^\circ$ (with spatial bin size of 0.02°). We used slices of these events maps, centred on the different sources of interest, as templates for our simulations.

Our sources of interest are 12 firmly identified PWNe among the HGPS sources, consisting of ten extended and two compatible with being point-like (HESS J1747-281 and HESS J1818-154) PWNe. These sources and their properties are listed in Table 1 (see also tables 3, 10, and 11 in H. E. S. S. Collaboration 2018). We have restricted the source library to these firmly identified PWNe for two practical reasons. First, the sample is sufficient in size to simulate a significant amount of different PWNe combined in pairs (i.e. artificially confused) in a limited computational time. Secondly, to facilitate the analysis of the performance of the simulations and the interpretation of the results, we prefer to have the input sources as best characterized as possible from the outset.

All source templates depict a square region of 3° side in 151×151 spatial bins. The bin size of the templates, i.e. a box of size 0.02°, is comparable to the best angular resolution predicted for the CTA performance[4]. Other H.E.S.S. sources, not of our interest but lying in the field of view of the templates, were masked and excluded from

the simulations. Finally, to account for various orientations of the sources, we appended ten different rotations (in steps of 36°) of each template to the library of source templates.

It is interesting to quantify the degree of similarity between the templates and rotations of themselves. The latter gives us an estimation of how well we can describe the source morphology. For instance, matching particular morphological features of the templates with data must be ruled out if we cannot determine their orientation compared to data. We chose Pearson's correlation coefficient ($C_{i,j}$), defined by equation (1), as a measurement of the morphological autocorrelation degree of the templates. We calculated the mean and deviation of the statistic from nine different rotations of each template in steps of 36°. Table 2 summarizes the results. The coefficient is computed as:

$$C_{i,j} = \frac{\sum_m \sum_n \left(Z_{i;\,m,n} - \bar{Z}_i \right) \times \left(Z_{j;\,m,n} - \bar{Z}_j \right)}{\sqrt{\left(\sum_m \sum_n Z_{i;\,m,n} - \bar{Z}_i \right)^2 \left(\sum_m \sum_n Z_{j;\,m,n} - \bar{Z}_j \right)^2}},\quad (1)$$

where i and j stand for the ith and jth rotation of one template, $Z_{m,n}$ indicates spatial bins in image coordinates, for templates with dimensions $M \times N$, and $\bar{Z}_i$ corresponds to the ith template's average. Pearson's correlation coefficient of two nearly identical templates approaches $C_{i,j} = 1$.

2.2 Simulation tools

The simulations we performed were implemented with CTOOLS (version 1.7.4) software package. CTOOLS[3] (Knödlseder et al. 2016) has been developed for the scientific analysis of data obtained with the existing and future Cherenkov telescopes, such as H.E.S.S., MAGIC, VERITAS, and CTA. It is based on GAMMALIB (consisting of a C++ library and a PYTHON module; Knödlseder et al. 2016), which provides a framework for the analysis of astronomical gamma-ray data.

The simulations were carried out with the CTOBSSIM tool, which creates simulated events lists using the instrument characteristics (specified by the Instrument Response Functions, IRFs) and an input model comprising a list of sources with specific spectral and spatial models. The simulated events lists can be represented in a sky map with the tool CTSKYMAP (see e.g. Fig. 1). By default, the sky maps of the simulated observations are shown in celestial coordinates with Cartesian projection. The exposure time of each observation is 25 h. The simulations were analysed with the tool CTLIKE, performing an unbinned maximum-likelihood fit to the data. The following parameters of the fitting model were left free to vary: (1) all parameters of the source power-law ($\mathrm{d}N/\mathrm{d}E = N_0 \times (E/E_0)^{-\Gamma}$) or exponentially cut-off power-law ($\mathrm{d}N/\mathrm{d}E = N_0 \times (E/E_0)^{-\Gamma} \times \exp(-E/E_{\mathrm{cutoff}})$) spectra, except the reference energy (E_0), (2) the norm and tilt of the background spectral model, and (3) the position and width (σ) of the sources with the spatial model corresponding to a radial Gaussian. The background component, provided by the CTA IRFs, is modelled by a template predicting the background rates as a function of position in the field of view and energy. The model is multiplied by a spectral power-law component, such that the energy distribution of the background is determined by fitting its amplitude and slope (tilt). The spatial model can be also specified using a template sky map containing the source of interest and describing an

[3]http://cta.irap.omp.eu/ctools/about.html

Table 1. Summary of firmly identified extended PWNe among the HGPS sources. The columns respectively list the following characteristics as reported in HGPS: name, related object, galactic longitude (l) and latitude (b), source extension (σ) in Gaussian sigmas, square root of the test statistic (TS), the parameters of the spectral model, as well as the integral flux over 1 TeV[a].

Name	Object	l (deg)	b (deg)	Size (σ) (deg)	$\sqrt{TS}$	$N_0 \times 10^{-12}$ (cm^{-2} s^{-1} TeV^{-1})	E_0 (TeV)	Γ	$E_{\rm cutoff}$ (TeV)	$F_{>1\rm TeV} \times 10^{-12}$ (cm^{-2} s^{-1})
HESS J0835-455	Vela X	263.96	-3.05	0.58 ± 0.052	39.4	6.41 ± 0.33	1.70	1.35 ± 0.08	12.3 ± 1.7	17.43 ± 1.40
HESS J1303-631	G304.10-0.24	304.24	-0.35	0.18 ± 0.015	54.5	7.45 ± 0.24	0.95	2.04 ± 0.06	15.12 ± 3.2	5.21 ± 0.35
HESS J1356-645	G309.92-2.51	309.79	-2.50	0.23 ± 0.020	17.3	0.57 ± 0.05	2.74	2.20 ± 0.08	–	4.39 ± 0.39
HESS J1418-609	G313.32 + 0.13	313.24	0.14	0.11 ± 0.011	21.9	0.83 ± 0.05	1.87	2.26 ± 0.05	–	2.69 ± 0.15
HESS J1420-607	G313.54 + 0.23	313.58	0.27	0.08 ± 0.006	27.6	0.84 ± 0.04	1.87	2.20 ± 0.05	–	2.77 ± 0.15
HESS J1514-591	MSH 15-52	320.32	-1.19	0.14 ± 0.026	42.0	7.95 ± 0.31	0.95	2.05 ± 0.06	19.20 ± 5.0	5.72 ± 0.42
HESS J1554-550	G327.15-1.04	327.16	-1.08	0.02 ± 0.009	9.1	0.058 ± 0.011	2.26	2.19 ± 0.17	–	0.29 ± 0.06
HESS J1825-137	G18.00-0.69	17.53	-0.62	0.46 ± 0.032	76.5	69.5 ± 2.9	0.65	2.15 ± 0.06	13.57 ± 3.9	19.15 ± 1.85
HESS J1837-069	G25.24-0.19	25.15	-0.09	0.36 ± 0.031	41.5	20.0 ± 0.7	0.95	2.54 ± 0.04	–	11.55 ± 0.49
HESS J1849-000	G32.64 + 0.53	32.61	0.53	0.09 ± 0.015	9.1	0.077 ± 0.010	2.74	1.97 ± 0.09	–	0.58 ± 0.07
HESS J1747-281	G0.87 + 0.08	0.87	0.08	Point-like	-	0.84 ± 0.13	1.0	2.4 ± 0.11	–	0.60 ± 0.13
HESS J1818-154	G15.4 + 0.1	15.41	0.16	Point-like	5.6	0.11 ± 0.02	1.54	2.21 ± 0.15	–	0.7 ± 0.2

Note. [a] See tables 3, 10, and 11 in H. E. S. S. Collaboration (2018) for further detail.

Table 2. Pearson's correlation coefficient ($C_{i,j}$) computed for each template compared to a rotation of itself.

Name	$\langle C_{i,j} \rangle$
HESS J0835-455	0.72 ± 0.03
HESS J1303-631	0.88 ± 0.01
HESS J1356-645	0.44 ± 0.03
HESS J1418-609	0.58 ± 0.02
HESS J1420-607	0.71 ± 0.01
HESS J1514-591	0.89 ± 0.02
HESS J1554-550	0.87 ± 0.02
HESS J1825-137	0.75 ± 0.06
HESS J1837-069	0.67 ± 0.03
HESS J1849-000	0.77 ± 0.03

arbitrary intensity distribution. In the latter case, the spatial model does not include free parameters.

The characteristics of the CTA performance,[4] extensively studied with Monte Carlo (MC) simulations of the CTA instrument (Acharyya et al. 2019) based on the CORSIKA air shower code (Heck et al. 1998) and telescope simulation tool SIM_TELARRAY (Bernlöhr 2008), are provided by the Instrument Response Functions (version PROD5 v0.1, Observatory & Consortium 2021). The IRFs were calculated for the planned southern and northern arrays and for different sub-arrays of telescopes that observe an object at three zenith angles (i.e. 20°, 40°, and 60°). Different analysis cuts are applied for each IRF, considering observation times of 0.5, 5, and 50 h. The PROD5 v0.1 version of the IRFs we employed assumes the CTA arrays in the dubbed Alpha configuration, i.e. accounting for 4 LSTs and 9 MSTs in the northern array and 14 MSTs and 37 SSTs in the southern one (spread over an area approximately of 0.25 and 3 km², respectively). We referred the results to the northern and southern arrays at 20° zenith angle with the IRFs referenced to 50 h.

2.3 Simulating CTA observations

We simulated a 25-h observation of each source, isolated and in different orientations, with the northern and southern arrays of CTA. We consider 25 h to be a reasonable estimate of the average

[4]https://www.cta-observatory.org/science/ctao-performance/

observation time for the selected targets during the CTA GPS since 1620 h of total observation time (1020 h with the southern array and 600 h with the northern one) are requested for the GPS during a 10 yr programme (see Cherenkov Telescope Array Consortium 2019). In comparison, the average livetime for the H.E.S.S. observations of the PWNe listed in Table 1 is ∼50 h, comprising the HGPS 2864 h of total observation time (H. E. S. S. Collaboration 2018). We performed all simulations with an energy threshold of 500 GeV (corresponding approximately to those of the templates, see table 11 in H. E. S. S. Collaboration 2018), taking the input positions and spectral models of the sources from the HGPS source catalogue. Fig. 1, e.g. shows some simulations corresponding to Vela X and MSH 15-52 in their unaltered orientation compared to the templates of the sources (in the top panels).

The sources were next artificially confused in pairs, resulting in 66 ($= 12 \times 11/2$) possible pairings for the chosen sample, not accounting for different projected separations or rotations. First, we place two templates in the same position, with one of them taken as a reference. We relocate a template by simply modifying its central bin's reference position (preserving the bin size). Next, we introduce between the sources a random separation retrieved from a Gaussian distribution centred in 0.5° with 0.25° of full-width at half-maximum (FWHM).

We obtained the mean separation between the sources (i.e. $d_{\rm PWN} \approx 0.5°$) from an estimation of the projected density of PWNe ($\rho_{\rm PWN}$) in a central region of the Galaxy defined by $|l| < 30°$ and $|b| < 0.5°$. We considered; $\rho_{\rm PWN} = N_{\rm PWNe}/A = 1/(\pi d_{\rm PWN}^2/4)$, where A corresponds to the area (i.e. 60 deg²) and $N_{\rm PWNe}$ to the number of sources therein. The latter was computed from a Galactic source distribution model (Renaud & CTA Consortium 2011; Fiori et al. 2022) that has been used in previous works to evaluate the source confusion problem in the CTA Galactic survey. We obtained from the same $N_{\rm PWNe} \sim 188$ (or $\rho_{\rm PWN} = 3.1$ sources per square degree in the cited region) at a sensitivity level of 3 mCrab (i.e. 5.24×10^{-14} cm^{-2} s^{-1} of flux above 1 TeV; Dubus et al. 2013).

In a former, limited study of confusion, see Cherenkov Telescope Array Consortium (2019), the exercise evaluated whether two sources were positionally coincident within a certain radius defined through the CTA PSF. In particular, the definition adopted posed that a position in the sky was confused if there was more than one simulated source within a radius of 1.3 times the CTA's angular resolution. In these studies, the sources were taken from
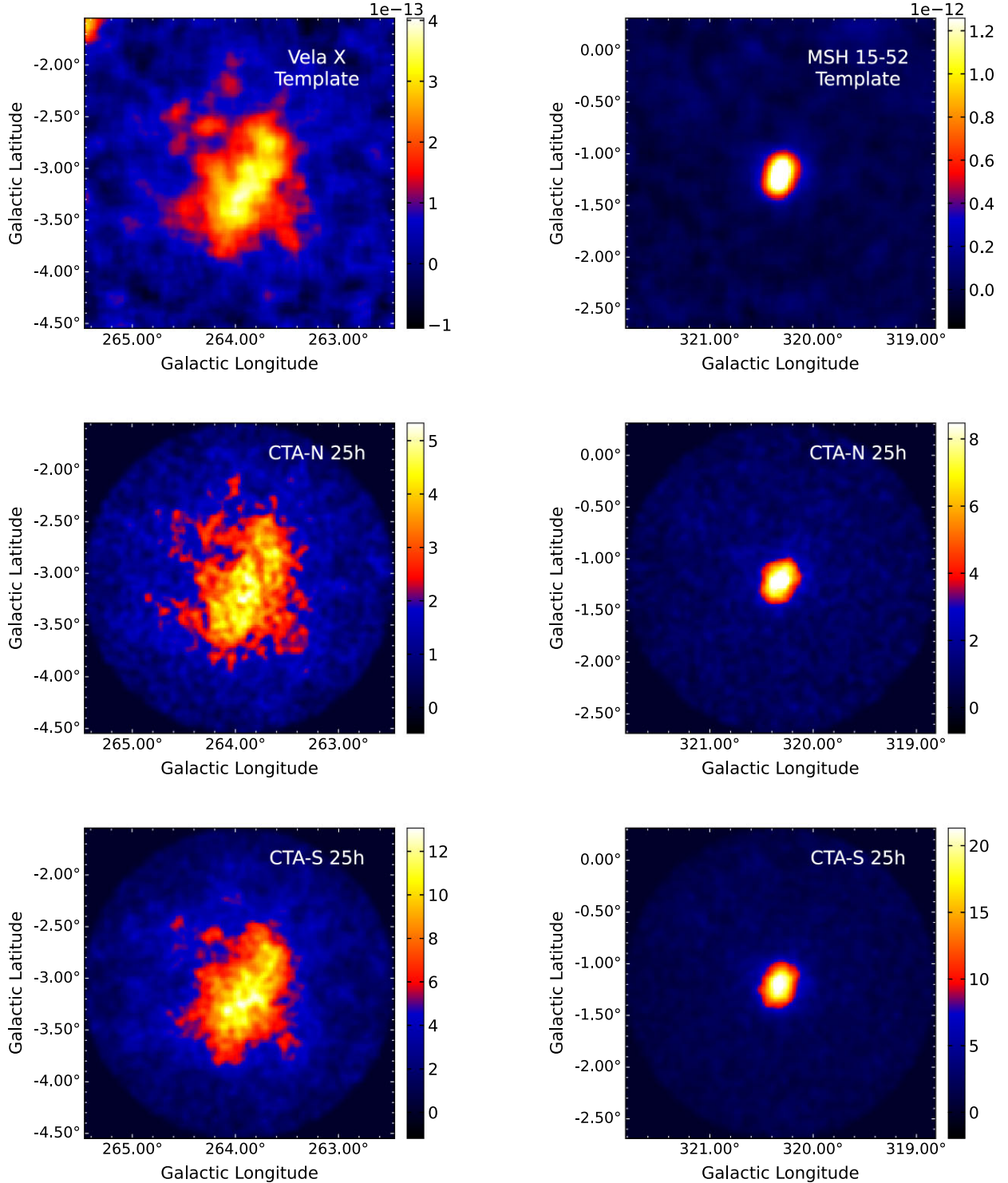
Figure 1. The top panels depict the morphological templates of Vela X (left-hand panel) and MSH 15-52 (right-hand panel), obtained from the H.E.S.S. Galactic plane Survey. In the central and lower panels: the sky maps of Vela X and MSH 15-52 for the 25-h simulated observations with the northern (CTA-N) and southern (CTA-S) arrays of CTA above 0.5 TeV. The plots have been smoothed with a Gaussian kernel of $\sigma \approx 0.04°$. The colour bars located at the side of the templates are in units of integral flux (cm^{-2} s^{-1}) over 1 TeV (normalized by the factor shown on top of the bar), while those at the side of the simulated CTA sky maps show the number of events (counts) per pixel of $0.02°$ without background subtraction.

different extrapolations of the total source count (as a function of flux, i.e. $\log N - \log S$ diagram), sizes, and spectral indices distributions that were assumed to be consistent with existing data. A specific spatial distribution of sources around the Galactic Centre

was assumed, with no diffuse emission except for the Galactic Centre ridge (see Section 6.4.2 of the cited work). The main limitations to consider were the unknown shapes of the sources, the undetermined level of diffuse emission, the high source density in the inner

Galaxy, and the dependence of source identification on the analysis methods.

Unlike in Cherenkov Telescope Array Consortium (2019), we established the following more general criterion: two sources are strictly confused when $\sigma_1 + \sigma_2 > d$, where σ_1 and σ_2 are the Gaussian widths of the sources (see Table 1) and d the projected separation between their centroids. We set $\sigma = 0.02°$ for the point-like sources. CTA's predicted angular resolution (68 per cent PSF containment radius) at a few TeVs is smaller than $0.05°$ for both the northern and southern arrays. It is then considerably smaller than the average projected separation of the simulations performed, i.e. $0.5°$, and small compared to the Gaussian size of most sources in the library. For this reason, we can neglect the CTA PSF in the confusion criterion above. In contrast to the initial studies on source confusion cited in Cherenkov Telescope Array Consortium (2019), the CTA's identification capabilities in our simulations are then limited by noise rather than by the instrument's PSF.

Finally, we rotate clockwise the template previously shifted by an angle corresponding to a random multiple of 36° (including 360°) to explore different relative orientations among the templates. Fig. 2 depicts three simulations of Vela X and MSH 15-52 artificially confused with random separations taken from the probability distribution discussed above and different orientations of MSH 15-52 as an example. In the lower panel, the sources are not strictly confused according to the separation imposed (0.72°). The panels are centred at the position of the Vela X template.

To restrict the computational time, we limited the simulations to two different configurations for each possible pair of confused PWNe out of the source library, i.e. each configuration consisting of an arbitrary projected separation and relative orientation. The 66 possible pairings we have, with two CTA arrays and using two different configurations for each pair of confused sources, resulted in 264 simulations. The total census of templates involved in these simulations amounts to 252 templates, i.e. 120 templates depicting one source (twelve source templates in ten different orientations) and 132 templates of artificially confused sources (66 pairings in two different configurations each).

2.4 Analysis of the simulations

We fitted the different source templates (i.e. considering the various orientations) and a Gaussian source to the simulations of the sources as isolated. We dubbed the cited hypotheses $H_{\text{Temp.i},\alpha}$ – where the first subindex identifies the template and α the rotation angle – and H_{Gauss}, respectively. The source detection significance (in Gaussian σ) was approximated as the square root of the Test Statistic ($\sqrt{\text{TS}}$). The Test Statistic is defined from the maximum log-likelihood value obtained when fitting the source (together with the background) to the data ($\ln L$) and the same if only fitting the background model ($\ln L_0$), as $\text{TS} = 2 \times \ln(L/L_0)$. The TS was then used to compare the goodness of fit among the different hypotheses employed to model each observation simulation.

To analyse the simulations of confused sources performed, we considered different hypotheses. First, we fitted the following models to each simulation:

(i) A source described by a template resulting from two confused sources ($H_{\text{Conf.}}$) and exponentially cutoff power-law spectrum. This hypothesis accounts for all pairings of PWNe from the library in the two configurations considered, i.e. $H_{\text{Conf.}}$ comprises all 132 templates used for simulating artificially confused sources (see Section 2.3). Conceptually, we cannot interpret the latter as a proper 3D fit to

the simulated data since the sources have different spectra and each confused template fitted to the simulations has a unique spectrum. The $H_{\text{Conf.}}$ hypothesis, however, can still represent quantitatively well a given observation simulation.

(ii) A source with a radial Gaussian spatial model and exponentially cutoff power-law spectrum (i.e. H_{Gauss}).

(iii) The source templates in the library with all their corresponding rotations ($H_{\text{Temp.i},\alpha}$). We dubbed the particular cases in which the fitting template is one of those involved in the simulation as $H_{\text{Temp.1},\alpha}$ and $H_{\text{Temp.2},\alpha}$. We fitted the one-source templates considering the spectral shapes listed in Table 1.

None of the cited hypotheses, fitted to a given simulation of confused sources, can reproduce the input model. However, they can represent (or fit) well the simulated observation. This hence allows us to probe the probability for two confused sources to be well-described by only one source.

We next compared the goodness-of-fit corresponding to the different hypotheses with the value of the Test Statistic (TS). Note that

$$\text{TS}_{H_i} - \text{TS}_{H_j} = 2 \times \ln(L_i/L_j), \quad (2)$$

where $\ln(L_i)$ is the maximum log-likelihood value corresponding to the hypothesis H_i. Therefore, if we define:

$$\Delta\text{TS} = \max(\text{TS}_{\text{Conf.}}) - \max\left(\text{TS}_{\text{Gauss}}, \text{TS}_{\text{Temp. 1},\alpha},\right.$$
$$\left.\text{TS}_{\text{Temp. 2},\alpha}, \text{TS}_{\text{Temp.i},\alpha}\right) \quad (3)$$

we can conclude that a positive and large ΔTS for a given observational data set translates into better prospects for resolving the two (confused) sources. For example, the two cases shown in Table 3, satisfy $\Delta\text{TS} > 0$ for both the CTA northern and southern arrays. In this case, the MSH 15-52 template (i.e. $H_{\text{Temp.2},\alpha}$ hypothesis in equation 3), could not explain the simulated data; $\text{TS}_{\text{Temp.2},\alpha} \approx 0$. Alternatively, we can compare the different hypotheses fitted to the simulated data using Akaike's Information Criterion (AIC; Akaike 1973, 1974) for not nested models. However, we reached the same conclusions after analysing the simulations as if using equation (3) (see Appendix A for a detailed comparison of AIC and TS statistics).

When $\Delta\text{TS} > 0$ happens, we try to decompose the best-fitting template corresponding to the hypothesis $H_{\text{Conf.}}$ in the two templates taken from the library, assigning each one to a different source. For this purpose, we perform the unbinned maximum-likelihood fit to the given simulation considering the two sources cross-matched with the templates library, each with its corresponding spectrum and source template as the spatial model (plus the background, H_{2src}). When fitting the latter hypothesis, we can indeed retrieve the simulated model provided a correct identification of the sources from the cited cross-match. As mentioned in Section 2.2, the free parameters of the fit are the ones of the spectral models (except the reference energies) and those of the background. The TS of the ith source (where $i \in \{1, 2\}$) is designated as $\text{TS}_{\text{2src}, i}$. Note that one of the sources has been shifted and rotated. Hence, the position and orientation used to model the latter source are inherited (fixed) from the best-fitting template under the $H_{\text{Conf.}}$ hypothesis.

3 RESULTS

3.1 Simulations of isolated sources

We analysed the simulations of the sources listed in Table 1 (see Section 2.3), assuming that the sources are isolated and have different arbitrary orientations. For this purpose, we fitted all hypotheses
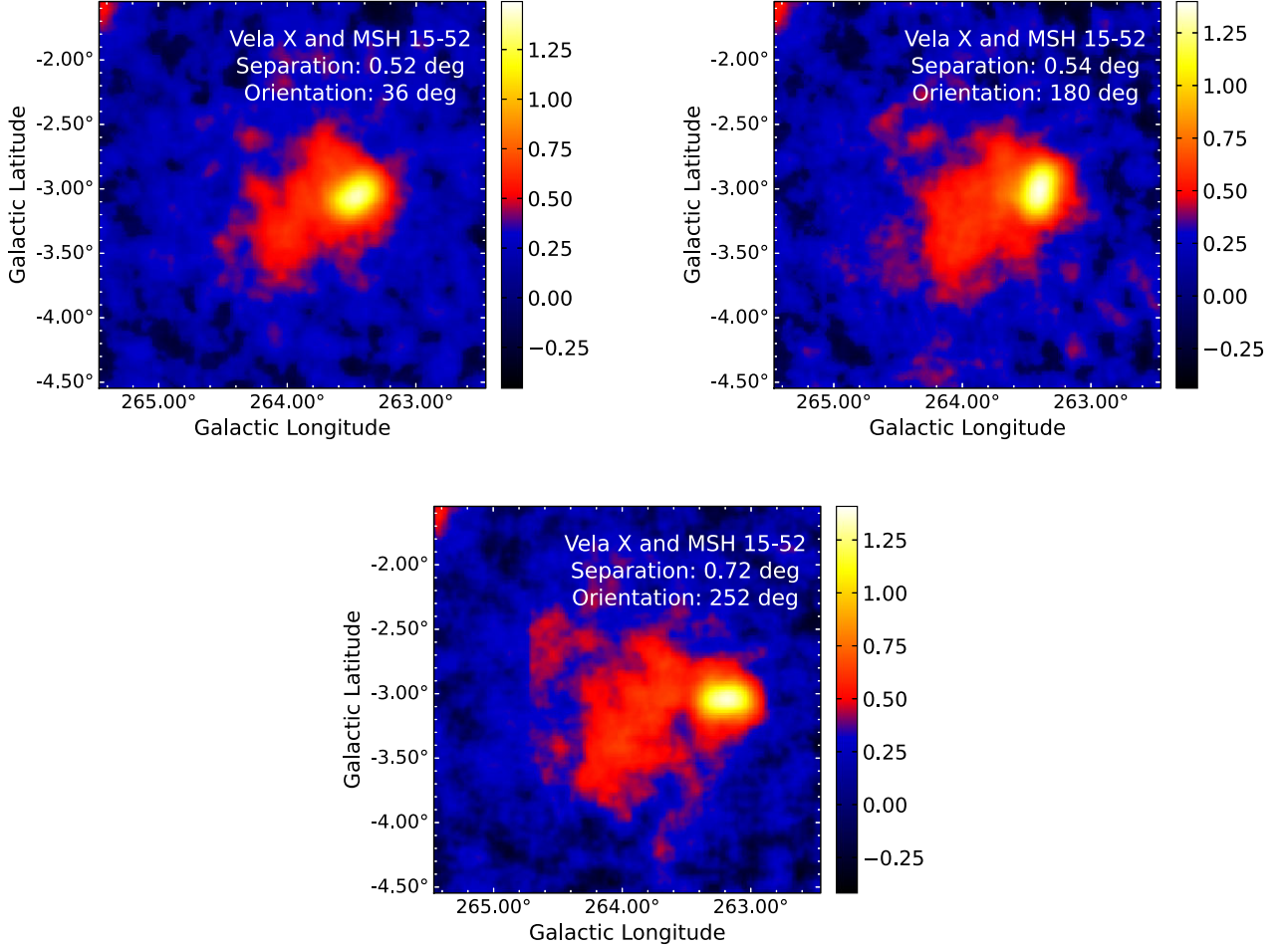
Figure 2. The morphological templates of Vela X and MSH 15-52 artificially confused, i.e. separated by 0.52° (top left-hand panel), 0.54° (top right-hand panel), and 0.72° (bottom) with MSH 15-52 rotated clockwise by 36°, 180°, and 252°, respectively, compared to the orientation from H.E.S.S. data. The colour bars at the side of the templates are in units of the square root of the integral flux over 1 TeV ($\times 10^6$ cm^{-2} s^{-1}), preserving the sign in each bin. We applied this transformation to the addition of the templates to better visualize the contribution of both sources. The plots have been smoothed with a small Gaussian kernel ($\sigma \sim 0.04°$).

Table 3. The square root of the TS resulting from fitting to the CTA simulated observations of MSH 15-52 and Vela X (artificially confused): (1) The Vela X template morphology (top left-hand panel in Fig. 1). (2) A Gaussian source model. (3) A source with the morphology of the MSH 15-52 and Vela X confused templates (see Fig. 2). The hypotheses listed correspond to: $H_{\text{Temp.1},\alpha = 0°}$, H_{Gauss}, and $H_{\text{Conf.}}$ in equation (3), respectively.

Parameter	CTA Array	Vela-X	Gaussian	Two sources
0.52° separation				
36° orientation				
$\sqrt{\text{TS}}$	North	75.5	81.0	90.9
$\sqrt{\text{TS}}$	South	121.0	127.3	139.7
0.54° separation				
180° orientation				
$\sqrt{\text{TS}}$	North	72.5	76.8	83.7
$\sqrt{\text{TS}}$	South	116.2	122.7	131.4

(see Section 2.4) regarding an isolated source to each simulation. As an example, Table 4 summarizes the best-fitting model for the simulations of Vela X depicted in Fig. 1. The results of fitting the Vela X template morphology (left-hand panel in the cited figure) and a Gaussian model to the simulated data are summarized in the top and bottom parts of the table, respectively. We can conclude that the fits successfully recover the source's characteristics.

The outcome of the simulations indicates a remarkable capability to match a simulation with both the correct input template and its proper orientation. In more than 95 per cent of the simulations, the best-fitted model (with maximum TS) identified the input template with the correct orientation. In addition, in more than 90 per cent of cases the best-fitting template model qualitatively improved a Gaussian source fit: $\text{TS}_{\text{Template}} - \text{TS}_{\text{Gauss}} \gtrsim 25$. Furthermore, more than 80 per cent of the simulations of the faintest and less extended sources in the library (i.e. $F_{>1\text{TeV}} < 10^{-12}$ cm^2 s^{-1} and $\sigma < 0.1°$) are correctly matched with the input template and orientation. These simulations are also seemingly best represented by the template compared to a Gaussian source. The left-hand panel of Fig. B1 depicts some examples of how the detection significance varies with the orientation of the template fitted to a simulation. None of the simulations concerning an isolated nebula resulted best represented by a confused template compared to the best-fitting hypothesis regarding one source, i.e. equation (3) satisfied $\Delta\text{TS} < 0$ for all simulations of isolated sources.

Table 4. The best-fitting parameters to the CTA 25-h simulations of Vela X (central and bottom left-hand panels of Fig. 1) compared to the source's input model.

Parameter	Fixed?	Model	CTA-N result	CTA-S result
Template fit:				
$\sqrt{\text{TS}}$ (Vela X template)			102.3	167.1
Spectrum results				
$N_0(10^{-12}\text{cm}^{-2}\text{s}^{-1}\text{TeV}^{-1})$	No	6.41	6.42 ± 0.12	6.33 ± 0.08
$E_0(\text{TeV})$	Yes	1.7	1.7	1.7
Index	No	1.35	1.37 ± 0.03	1.33 ± 0.02
$E_{\text{cutoff}}(\text{TeV})$	No	12.3	12.5 ± 0.6	12.1 ± 0.3
Gaussian fit				
Ra	No	128.887	128.812 ± 0.011	128.831 ± 0.006
Dec	No	-45.659	-45.662 ± 0.008	-45.662 ± 0.004
Size (Gaussian)	No	0.58	0.542 ± 0.006	0.541 ± 0.004
$\sqrt{\text{TS}}$ (Gaussian)			95.5	152.9

3.2 Simulations of confused sources

Fig. 3 summarizes the results. The ratios are computed from the 264 simulations with the 12 PWNe listed in Table 1, with an uncertainty estimated of $\sim [1–2]$ per cent. We calculated the latter error through bootstrapping with the method further explained in Section 3.3 and applied in Fig. 5. Note that Table 3 and Fig. 2 already exemplified a particular result among those accounted for in Fig. 3. Since the separations among each simulated pair of sources are obtained randomly from a probability distribution, it is not guaranteed that the confusion criterion holds for each simulation. Secondly, independently of whether the sources are strictly confused or not, we can classify them according to the result of equation (3), i.e. in simulations best fitted to the hypothesis $H_{\text{Conf.}}$ (i.e. $\Delta\text{TS} > 0$) or any other hypothesis considered ($\Delta\text{TS} < 0$). Note that, in most simulations, the sources are not strictly confused (see the dark and light green sectors in the left-hand chart of Fig. 3), although this does not exclude one of the sources lying on top of the extended emission of the other. The latter is possible given the angular resolution of VHE gamma-ray telescopes, of few arcminutes, and the various extensions of the sources considered, from point-like to very extended ones, as e.g. HESS J0835-455.

All sources in the library would be detected above 5σ after 25 h, both with the CTA northern and southern arrays when isolated. The latter was expected, since H.E.S.S. detected all of them in observation times of $\sim[20–50]$ h. In this case, the input model from the simulation can be retrieved with small statistical errors, see e.g. Table 4. However, detecting all sources in the library when artificially placed on top of (or very close to) each other is no longer guaranteed. Note in Table 1 that the integrated TeV flux of any two sources taken from the library may differ by a factor larger than fifty.

From the simulations performed, approximately 30 per cent were strictly confused. Only about 18 per cent of the simulations corresponded to sources strictly confused with the data best described by the $H_{\text{Conf.}}$ hypothesis. It is important to note that the population of point-like and/or dim unresolved nebulae in the source library crucially affects the latter proportions. For example, compare the right-hand chart of Fig. 3, in which the point-like sources together with the two dimmest extended nebulae (i.e. HESS J1554-550 and HESS J1849-000) were excluded of the source library, with the complete results of the simulations (at the left). Hence, the results obtained for our simulations are only valid if the source library is representative of the TeV population of Galactic PWNe.

Since the population of point-like and/or dim nebulae that CTA will detect is surely underestimated by our library – which is based

on H.E.S.S. data – we must regard the cases in which the source confusion problem would be likely resolved by CTA (i.e. 18 per cent) as an optimistic prospect. We will discuss the issues brought by dim sources more in-depth below.

Once we probe the presence of two sources ($\Delta\text{TS} > 0$) in a simulation, we can perform a joint fit of both sources (each with its corresponding template from the library), retrieving the input model of the simulation with high precision (see Figs B2, B3, B4, B5, and B6). The sources not detected at 5σ significance in all simulations performed are only those in Table 1 with integral flux at energies above 1 TeV smaller than $\sim 10^{-12}$ cm^{-2} s^{-1} (see Fig. B6). However, in some cases, these dim and small sources could be detected in 25 h, particularly when not artificially confused with a bright (more extended) source.

The effect of changing the CTA array employed in the simulations on to equation (3) is exemplified in the bottom panel of Fig. 4 (with vertical arrows). The simulations indicate that, in general, the detection significance of the confused sources hypothesis compared to the isolated source one, i.e. $\sqrt{|\Delta\text{TS}|}$ with ΔTS computed according to equation (3), increases approximately linearly with the projected separation between the sources. Fig. B7 illustrates the general trend mentioned. However, the same is not apparent in the bottom panel of Fig. 4, which only portrays two different projected separations.

3.3 Issues brought by dim sources

Fig. 4 shows the results of applying equation (3) to the observation simulations. The top left-hand panel corresponds to the northern CTA array and the right to the southern one. The simulations best represented by the $H_{\text{Conf.}}$ hypothesis have a positive Y-axis (i.e. light colours in Fig. 3), while the sources participating in a simulation are not strictly confused if $\sigma_1 + \sigma_2 < d$, i.e. at the left of the vertical line in Fig. 4 (in green tonalities in Fig. 3). We are thus most interested in the simulations located in the grey shaded area marked. As we expected, there are simulations that satisfy the confusion criterion; $\sigma_1 + \sigma_2 > d$, but are best fitted to one particular library template (either rotated or not, i.e. $\Delta\text{TS} < 0$), particularly if the confusion is strong, i.e. $\sigma_1 + \sigma_2 >> d$.

It is noticeable from comparing the top left-hand and right-hand panels in Fig. 4 that considering the southern array instead of the northern one for the same input configuration increases the source's significance obtained in the best-fitting hypothesis (whatever the same is). Typically, $\sqrt{\text{TS}}$ multiplies by a factor of $\sqrt{[2 − 3]}$ due
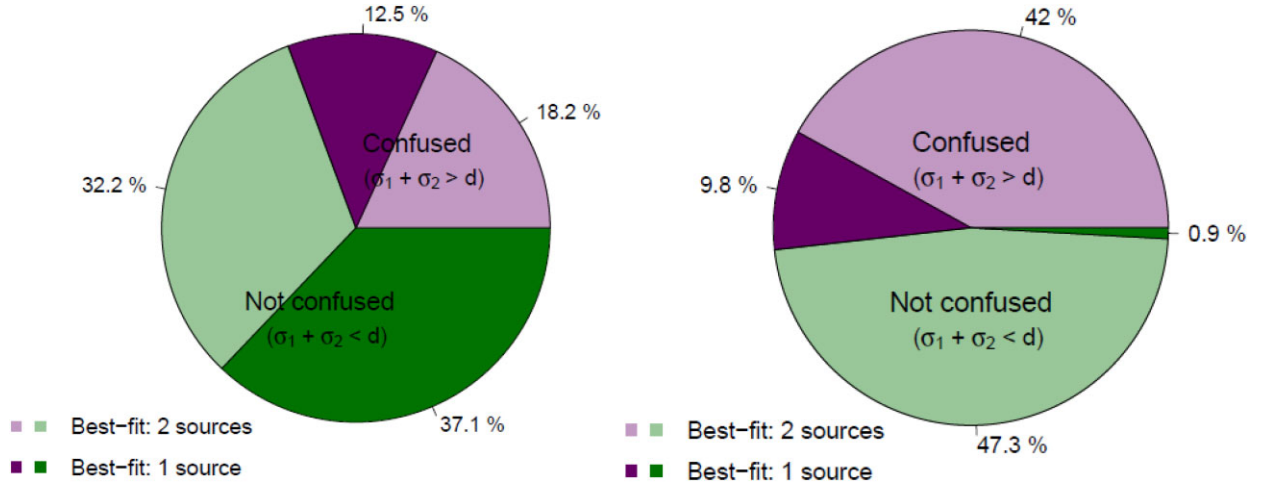
Figure 3. Distribution of the simulations of two sources at a certain distance according to whether the sources simulated are actually confused ($\sigma_1 + \sigma_2 > d$, in purple) or not (in green) and to whether the best-fitting model assumes one source (in dark tonalities) or two (in light tonalities). The right-hand panel excludes the simulations involving the two dimmest extended sources (HESS J1554-550 and HESS J1849-000) and those point-like.
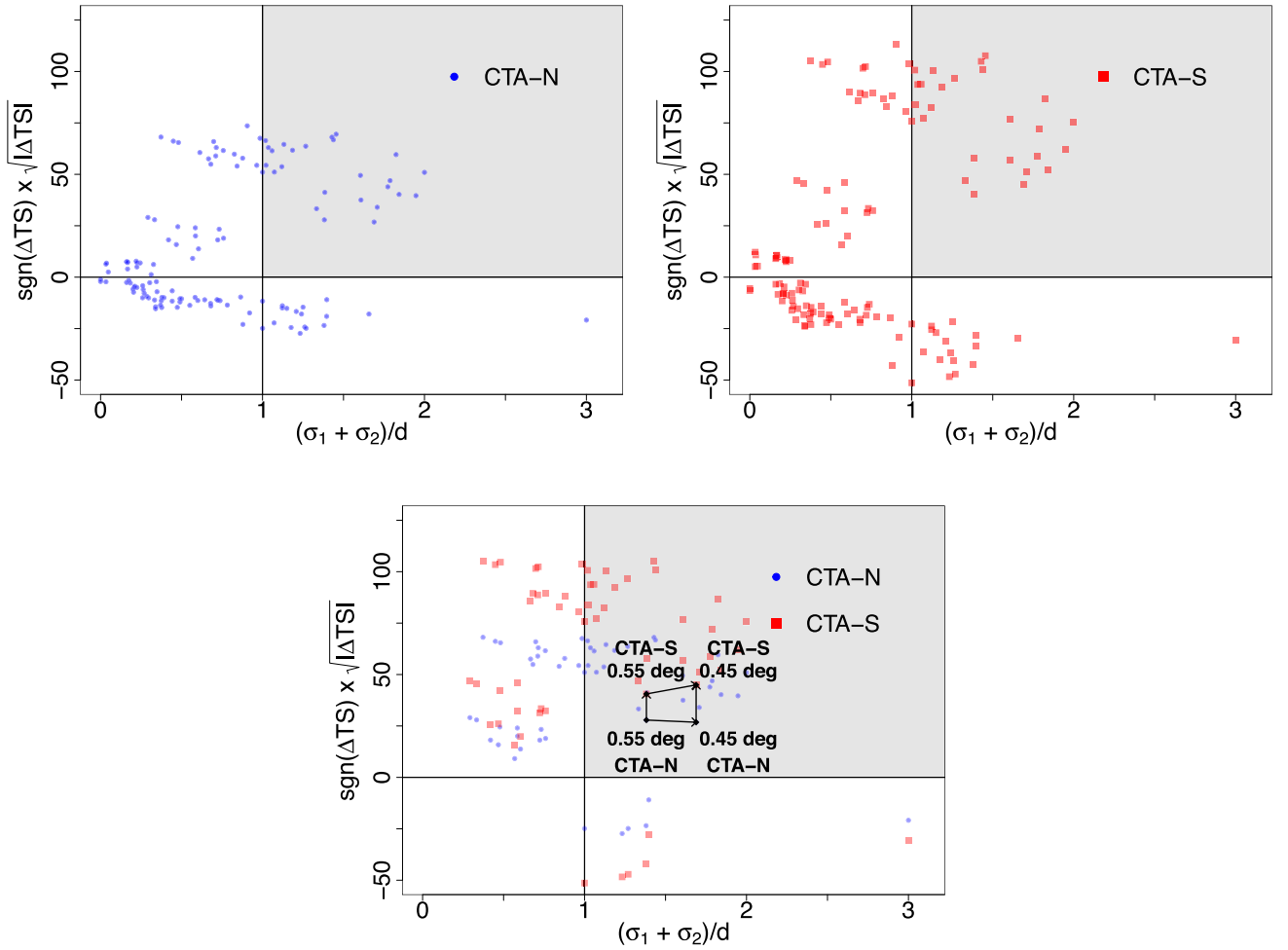

Figure 4. The square root of ΔTS in equation (3) (preserving the sign) is plotted for each pair of confused sources against the sum of their Gaussian sizes ($\sigma_1 + \sigma_2$, see the fifth column in Table 1) divided by the separation imposed. In the lower panel, we excluded the two dimmest extended sources (HESS J1554-550 and HESS J1849-000), and those point-like. The vertical black arrows exemplify how the results for the simulations (of a given pair of confused sources) move in the diagram when switching from the northern to the southern array. The horizontal arrows illustrate the effect of changing the separation between the sources.
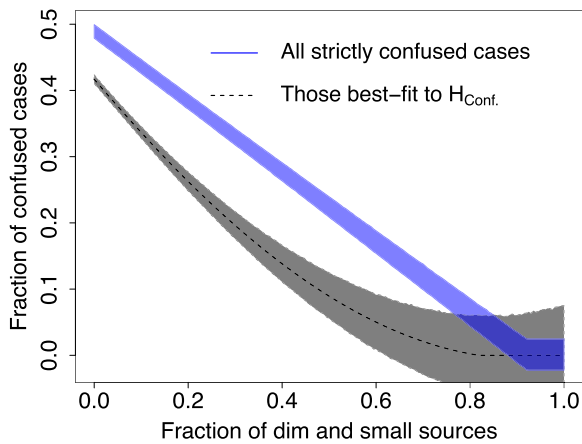
Figure 5. The fraction of simulations strictly confused and best fitted to the $H_{\rm Conf.}$ hypothesis (in black) and the total fraction of strictly confused ones (in blue) versus the fraction of faint and not very extended sources in the sample, i.e. with $S_{\rm b} < 10^{-10}$ cm^{-2} s^{-1} deg^{-2} and $\sigma < 0.1°$. The shadowed areas represent the 1-σ error regions.

to the improvement of sensitivity. However, the sign of ΔTS (from equation 3) is usually kept compared to that for the northern array.

A considerable number of observation simulations, as seen in the bottom left-hand quadrant in the panels at the top of Fig. 4, are best described by only one source, i.e. ΔTS < 0, despite not being strictly confused ($\sigma_1 + \sigma_2 < d$). Those cases are related only with four (out of the twelve) sources; HESS J1554-550, HESS J1849-000, HESS J1747-281, and HESS J1818-154, when placed close to a much more extended and/or luminous nebula (see the bottom panel of Fig. 4, which does not show the simulations related to the cited sources).

In the cases we referred to, the hypothesis regarding only the most extended and luminous among the sources represents the simulated data better than a combination of both source templates, even when according to the set up of the simulation, the sources are not strictly confused. It is apparent that, in these cases, simply perturbing the brightest source's spectrum may lead systematically to a better representation of data than if also slightly modifying its morphology (spatial template), resulting in ΔTS < 0. Fig. 3 may also be affected by this spectral effect. The confusion criterion does not likely hold in these cases due to the extension of one of the sources ($\sigma < 0.1$ or point-like). Consider HESS J1747-281 or HESS J1818-154, e.g. simulated at 0.65° of HESS J0835-455 (more luminous by a factor above 25). A joint fit of the two input source models to the observation simulations, in these cases, results in a significance achieved for the dimmest nebula well below 5σ. These simulations are noticeable in the top panels of Fig. B2, below the dashed line (TS < 25). It is not a surprising result that detecting a given source with CTA (or H.E.S.S.) may depend on the presence or not of bright nearby sources, mainly when the source is dim and/or small (or point-like).

The problem of addressing faint and small gamma-ray nebulae in the vicinity of bright and extended sources can only increase with the advent of CTA. Its improved sensitivity will translate into the detection of numerous dim sources. To account for this, we may, as a first approximation, extrapolate the results to arbitrary populations with different ratios of faint and small nebulae by resampling the simulations through a bootstrapping technique.

We carried out this extrapolation by extracting random samples from the simulations performed. Repeated elements in these samples were allowed, i.e. we subtract one randomly chosen template from

the whole sample at a time to generate new samples with varying percentages of dim sources. We used samples of various sizes from 5 to 45 simulations out of the 264 available, extracting 5×10^5 samples of each size. Next, we computed for each sample the ratio of simulations that are best fitted to the $H_{\rm Conf}$ hypothesis and the ratio of faint and small sources, treating each of the samples as a different population of sources.

We present the results in Fig. 5. The latter extrapolation, for example, resulted in $\sim$ 9 per cent of all simulations ($\sim$40 per cent of those strictly confused) best represented by the $H_{\rm Conf.}$ hypothesis (with an upper limit of 17 per cent at 95 per cent CL) for a fraction of dim and small sources of 0.5. The latter means that 50 per cent of PWNe in the input library would be similar in flux and extension to HESS J1554-550, HESS J1849-000, HESS J1747-281, or HESS J1818-154 ($S_{\rm b} < 10^{-10}$ cm^{-2} s^{-1} deg^{-2} and $\sigma < 0.1°$). In the cited case, about 22 per cent of the simulations strictly satisfied the confusion criterion. The simulations best fitted to the $H_{\rm Conf.}$ hypothesis (see black dashed line in Fig. 5), however, included wrong associations of templates to the simulated data. Out of the 264 simulations, the cases best represented by two confused sources in which one of the nebulae is incorrectly identified account for 8 per cent. In only one simulation both of the nebulae were misidentified.

The problem of wrongly associated templates will also increase with the addition of newly discovered faint and small nebulae (and the number of templates in the library). However, the extrapolations explained above to different populations of sources allowed us to predict the relative amount of possible mismatches with the template library in different scenarios. For a source population consisting of 50 per cent dim sources, the simulations best fitted by the $H_{\rm Conf.}$ hypothesis in which one of the nebulae is incorrectly identified were limited to $\sim$26 per cent (at a 95 per cent confidence level).

The difference between the grey and the blue shadows in Fig. 5 represents simulations that are not fitted to two confused sources, despite being confused in the input template. Note also in the figure that, as follows from the average separation between sources (i.e. 0.5°), the fraction of confused sources reaches zero when the fraction of dim and small or point-like sources is close to one. Hence, this study cannot characterize the source confusion problem regarding only these small and/or point-like sources.

3.4 Minimal autocorrelation degree to make two templates different

An expected degeneracy for α plus 180° in the best-fitting orientation of the templates, derived from the elliptical-like morphology of some of the nebulae templates employed, was observed (see Fig. B1). In some cases, the best-fitting rotation angle results close to the input one (black and red lines in the right-hand panel of Fig. B1). In others, it turns to a rotation angle separated by about 180° of the input value, e.g. see the blue line in the right-hand panel of the cited figure (where the input orientation is closer to the local maximum at 0°). The degeneracy observed in the best-fitting orientation is enhanced when the morphological autocorrelation degree of the simulated template approaches $C_{i,j} = 1$. We are then interested in the smallest variation of morphological autocorrelation degree that leads (for a given source) to a noticeable change in the source detection significance. This would be helpful, for instance, if the template library is built, e.g. from HD, MHD, or HD+B simulations. The smallest variation of $C_{i,j}$ detectable would be related, in this case, to the level of accuracy needed in the template simulations.

We performed simulations of a few nebulae among the population considered but with the source template blurred up to different
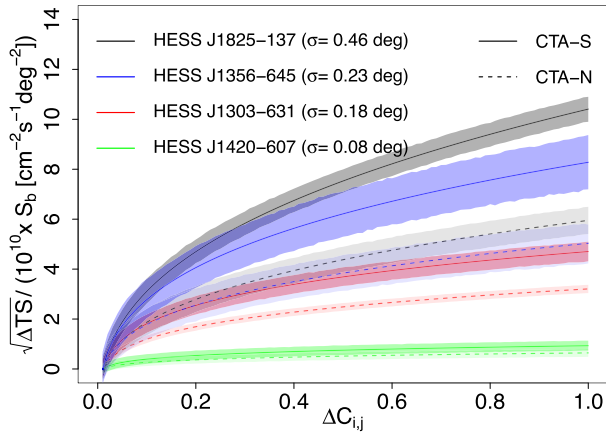
Figure 6. The change of the source detection significance when fitting a template to observation simulations performed with various degrees of blurring and those with the unaltered template, i.e. $\Delta\mathrm{TS} = \mathrm{TS}_{\mathrm{Blur.}} - \mathrm{TS}_{\mathrm{Unalt.}}$. We normalized $\sqrt{\Delta\mathrm{TS}}$ by the surface brightness of the source; $S_{\mathrm{b}} = 10.5, 6.2,$ 11.3, and 22.6 in $10^{-10}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{deg}^{-2}$, respectively, for each of the nebulae (in the same order of the legend). The Pearson's correlation coefficient ($C_{i,j}$, in equation 1) quantifies the similarity between the blurred and unaltered templates.

degrees. To blur the templates, we convolved its central region (a box of 1.5σ of side) with a function depending on an index parameter (β). This function takes the value $f(\mathrm{X}, \mathrm{Y}, \beta) = 2 - \max(|\mathrm{X} - \mathrm{X}_0|^\beta, |\mathrm{Y} - \mathrm{Y}_0|^\beta)/1.5\sigma$ inside the box, and $f(\mathrm{X}, \mathrm{Y}, \beta) = 1$ outside the box, where X and Y are pixel coordinates in the template and X_0 and Y_0 are referred to the template's centre. The index (ranging from 1 to 3) modulates the harshness of the blurring applied, corresponding the $\beta = 1$ case to the most similar template to the original. Fig. B8, provides an example of different degrees of blurring (quantified by the β parameter) applied to the HESS J1825-137 template. The Pearson's correlation coefficient between each blurred template and the unaltered one, where $\Delta C_{i,j} \approx 0$ for $\beta = 1$, is noted in each panel.

Next, we fitted the original (not altered) source template to the simulated data to probe the sensitivity of the simulations on the morphological autocorrelation degree. Fig. 6 shows how the detection significance changes with respect to the variation of the morphological autocorrelation degree for the blurred templates of different nebulae. In the figure, a larger $\Delta C_{i,j}$ indicates a more significant blurring applied. We normalized each curve by the surface brightness (S_b) of the corresponding nebula, derived from the template (in units of $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{deg}^{-2}$), to compensate for both the differences in flux and morphology of the sources at the same time. Using the surface brightness of particular PWNe, i.e. multiplying the Y-axis of Fig. 6 by the surface brightness values specified in the figure's caption, we obtained $\sqrt{\Delta\mathrm{TS}} \gtrsim 5$ for $C_{i,j} < 0.1$ in all the cases depicted. It means that our simulations are sensitive to such small variations of $C_{i,j}$ for PWNe as HESS J1825-137 or HESS J1303-631, which are among the rather luminous ones of our population. If the templates come from a theoretical prediction of a model, it would seem that two such templates should certainly be part of the library for $\Delta C_{i,j} > [0.05–0.1]$. Provided a given surface brightness, the size of the source is the most critical parameter, being the simulations of more extended sources more sensitive to deformations of the template (as highlighted in Fig. 6, see the sources extensions noted in Gaussian σ). Note that Pearson's coefficient depends on the template's binning. However, the latter compares well with the angular resolution

expected for CTA that will ultimately determine the template's bin size.

4 DISCUSSION: REAL DATA AND COMPUTATIONAL TIME

An analogous procedure using genuine CTA data would also start by fitting. On the one hand, we would fit a source with a model corresponding to each of the templates in the library considered on its own, taking either a set of rotations for each template or a rotation angle as a free parameter of the fit. Note that different rotations of a template can be treated, as we do, like any other template in the library (like a different PWN), which may simplify the fitting process. On the other hand, we would fit the same source with a model having different combinations of the templates of a putative library confused in pairs. The free parameters would be the separation between the templates and the relative orientation among the same. Next, similarly to equation (3), we would compare the TS corresponding to the best-fitting model among the hypotheses regarding two confused templates to that of the best-fitting model involving the template of only one source. In our simulations, the fit to a Gaussian source involves three degrees of freedom more (i.e. position and width) than a fit to a generic template. However, when fitting the templates from the library to actual CTA data, both hypotheses will have the same degrees of freedom since the templates will not maintain a reference position (accounting also for the rotation angle of the template).

We have accounted for different orientations of the simulated templates, handling the rotation of the templates straightforwardly by considering each rotated template a separate template of the library. However, we noted that although the simulations are sensitive to a transformation of the templates such as blurring, degradation, or rotation, most templates of the library have a high morphological autocorrelation degree. Fitting the orientation of the templates from the observation simulations is challenging due to their high autosimilarity, mainly because we observe a degeneracy between the best-fitting angle and rotations close to $180°$, as expected for elliptical-like morphologies. In addition, the absence of small-scale structures (below the HGPS angular resolution, i.e. $\sim 0.08°$) that CTA may resolve in the H.E.S.S. maps used as templates could prevent us from detecting differences involving small sources in these simulations.

As we used H.E.S.S. observational data, the diffuse interstellar background is included (implicitly) in the simulations. However, it is only so in an idealized case since CTA may estimate a different level for the diffuse interstellar emission than H.E.S.S. at the position of the simulated sources. Our simulations may be sensitive to such differences. We leave the analysis of this component more in-depth to future studies with real CTA data since CTA will map with unprecedented precision the large-scale diffuse emission at VHE gamma-rays.

Another limitation of this study is that we did not consider the physical distance of the sources fitted to the observation simulations. The same template can be artificially placed closer or further from a reference distance (r) by rescaling its flux (as $1/r^2$) and spatial axes (by $1/r$). The relative error for the latter approximation (from trigonometric arguments) is smaller than 0.01 per cent if $2r/D > 50$, where D is the physical diameter of the template. Fig. B9 central and right-hand panels, e.g. depict 25-h simulations of HESS J0835-455 with the CTA southern array rescaled at twice and four times the distance of the reference template, respectively. We can handle the rescaled templates and the rotated ones alike. However, the number

of templates in the library rapidly increases if considering different rotations at various distances for each source.

The computation time required to cross-match all the templates with an observation simulation will be crucial for future studies with a more comprehensive template source library. Hence, it is helpful to have a good estimate based on this study. We summarized the computation time to perform an observation simulation of two confused sources in Table B1 for different input configurations using a regular commercial laptop. To compare with Table B1, note that we used a FoV of 1.5° of radius for the simulations presented in this paper. This time does not include any likelihood fitting process. The combination of large observation times (more than 50 h) and low energy thresholds ($E_{th} < 100$ GeV) increases significantly (by a factor ten) the computational time of a single observation simulation. The number of simulations to perform at the end will also account for the two CTA arrays and multiple separations and orientations for each possible pair of sources from the sample, if not accounting also for various distances for each template. The cost in computing time of fitting different hypotheses to the observation simulations varies depending on the number of those considered and the number of free parameters involved. Still, the typical average computation time per fitted model (template) was [1–2] min. Hence, one observation simulation can be fitted against, e.g. 10^6 templates in one month if using a cluster with $\sim$30 CPUs. The total computation time for the simulations we present in this paper is about a month per 132 simulations on a regular commercial laptop. Note that the number of templates needed to account for all possible pairings from a source library grows with the number of sources (N) as $N \times (N - 1)/2$, not accounting for different orientations, separations, and/or distances.

Our simulations are based on CTOOLS, as explained in Section 2.2. However, two packages have been developed independently to implement the CTA Science Tools, i.e. CTOOLS and GAMMAPY (Deil et al. 2017; Nigro et al. 2019). The latter software was adopted as the Science Analysis Tools of the CTA by the CTA Observatory (CTAO), being this decision announced on 2021 June 1.[5] Both softwares were proven to be stable, providing nearly identical high-level analysis results (such as events sky maps or spectra) in different analyses (Mohrmann et al. 2019; Mestre et al. 2020). Hence, we do not expect significantly different results if implementing our simulation scheme in GAMMAPY. To perform simulations of source templates, either in CTOOLS or GAMMAPY, is nearly a linear operation if not accounting for background. Hence, we can combine the templates before the simulation or stack the simulations of two templates afterward (with no background), as is shown in Fig. B10. To perform this test of linearity, we assigned an arbitrary spectrum to the confused template simulated in the left-hand panel of Fig. B10. Next, we assigned to each template in the right-hand panel of the cited figure the spectrum used for the confused one (at the left) with the normalization parameter divided by two. Both panels can be thus directly compared.

5 CONCLUSIONS

This paper presents the most detailed quantitative study up-to-date on source confusion of extended sources and identification capabilities with CTA. The tools available to perform simulations of TeV sources with CTA allowed us to conduct simulations of two confused or closely located sources in a variety of configurations, involving different separations, relative orientations, flux levels, and

extensions. We showed how these simulations could be analysed through a direct comparison to a library of extended and/or point-like TeV source templates and that it is possible (in some cases at very high confidence) to associate a CTA simulation with a combination of two templates from the library placed at a small distance ($\sim$0.5°) of each other. We also demonstrated that if the latter association is reached and correct, the characteristics of both sources (despite lying one on top of the other) can be studied in detail. In these cases, we retrieve similar statistical errors to those obtained analysing the sources in the isolated scenario.

For applying the method to real data we note that we would first need to obtain a library of morphological templates to cross-match with observations, representative of the expected population of sources. Also, the templates used here are referred to a specific distance, and the source may present energy-dependent morphology in CTA data, which will further complicate the cross-matching of the simulations with the template library. We have not considered the source confusion cases regarding more than two sources, mainly affecting the central parts of the Galaxy, nor the contamination of other non-confused but nearby sources in the analysis of the simulations. Note that we have only considered non-variable sources. Additionally, the simulations are also limited by the uncertainty in the flux level of diffuse emission and the shapes and distribution of the extended sources. For instance, source confusion involving only small and/or point-like sources is marginally represented in our simulations.

Despite the cited limitations, we constrained the source confusion in CTA data, particularly regarding the future Galactic Plane Survey with CTA, limiting the amount of source confusion likely to be resolved (as a first approximation) to 18 per cent above 500 GeV with an upper limit of 23 per cent at 95 per cent CL (based on the currently available Galactic Plane Survey from the H.E.S.S. experiment). We also obtained an upper limit of 33 per cent (at 95 per cent CL) for the occurrence of strict source confusion, i.e. $\sigma_1 + \sigma_2 > d$. These numbers apply only if assuming that CTA will detect a similar population of PWNe to that detected by H.E.S.S.; which is optimistic for obvious reasons. The amount of presumably resolvable source confusion may be limited to only $\sim$10 per cent if more than half of the TeV PWNe population regarding the future CTA Galactic Plane Survey consists of faint and small sources (i.e. $F_{>1\text{TeV}} < 10^{-12}$ cm^2 s^{-1} and $\sigma < 0.1°$), which is likely a more realistic scenario.

The flawed cross-matches with the template library can be a problem of the approach. Approximately 8 per cent of the simulations performed, based on the Galactic Plane Survey from the H.E.S.S. experiment, presented a misidentification of sources. This problem can aggravate if accounting for a more realistic ratio of dim sources in the input library. For instance, this difficulty can appear in up to 26 per cent of (likely) resolved cases of source confusion for a 50 per cent (or higher) ratio of dim sources.

Even though, the general approach we present can provide handy information, facilitating the identification of numerous sources (as our results regarding isolated nebulae illustrate) and, to some extent, the resolution of source confusion cases. Even if achieving the latter is challenging according to our simulations, associating empirical or theoretical templates to CTA data could provide exhaustive information on multiple sources. Our approach matches the data with the knowledge applied to the template generation (handled by the user). Hence, given some data, it allows us to explore different hypotheses more in-depth than assuming a simple geometrical morphology for the sources as the usual Gaussian or point-like assumptions.

A 3D binned likelihood fitting approach (newly introduced for VHE gamma-ray data analysis) has recently been demonstrated as

---

[5]https://www.cta-observatory.org/ctao-adopts-the-gammapy-software-package-for-science-analysis/

an efficient technique to disentangle several regions with different spectral and morphological characteristics within a significantly extended emission (see Abdalla et al. 2021 and references therein). This kind of analysis, together with the (also 3D) approach we tested in this work, can potentially alleviate the source confusion furthermore.

To conclude, at the methodological level, we note that the general approach developed here can be applied too at other frequencies, allowing similar template comparisons with X-rays and radio data. Likewise, the methodology we present can be straightforwardly generalized to source confusion regarding other extended Galactic sources (e.g. between SNRs and PWNe) or extragalactic ones.

DATA AVAILABILITY

The empirical data used to gather the templates for the sources considered are public and available in the additional materials regarding the H.E.S.S. Galactic Plane Survey (H. E. S. S. Collaboration 2018), and were obtained from https://www.mpi-hd.mpg.de/hfm/HESS/hgps/. The CTA IRFs are public and available in https://www.cta-observatory.org/science/cta-performance/.

REFERENCES

Abdalla H. et al., 2021, A&A, 653, A152
Acharyya A. et al., 2019, Astropart. Phys., 111, 35
Aharonian F. et al., 2004, ApJ, 614, 897
Akaike H., 1973, in Csáki F., Petrov B. N., eds, 2nd International Symposium on Information Theory, 1973. Akademiai Kiado, Budapest, p. 267
Akaike H., 1974, IEEE Trans. Autom. Control, 19, 716
Albert A. et al., 2020, ApJ, 905, 76
Bernlöhr K., 2008, Astropart. Phys., 30, 149
Bernlöhr K. et al., 2003, Astropart. Phys., 20, 111
Cherenkov Telescope Array Consortium 2019, Science with the Cherenkov Telescope Array. World Scientific Publishing, Singapore
Deil C. et al., 2017, in Proc. Sci., 35th International Cosmic Ray Conference (ICRC2017). Sissa, Trieste, PoS#766
Dubus G. et al., 2013, Astropart. Phys., 43, 317
Fiori M., Olmi B., Amato E., Bandiera R., Bucciantini N., Zampieri L., Burtovoi A., 2022, MNRAS, 511, 1439
H. E. S. S. Collaboration 2018, A&A, 612, A1
Heck D., Knapp J., Capdevielle J. N., Schatz G., Thouw T., 1998, in GmbH Forschungszentrum Karlsruhe, ed., CORSIKA: a Monte Carlo Code to Simulate Extensive Air Showers. TIB Hannover, D-30167, Hannover
Knödlseder J. et al., 2016, A&A, 593, A1
Kolb C., Blondin J., Slane P., Temim T., 2017, ApJ, 844, 1
Mazin D., 2021, Proc. Sci., 395, 872
Mestre E., de Oña Wilhelmi E., Zanin R., Torres D. F., Tibaldo L., 2020, MNRAS, 492, 708
Mohrmann L., Specovius A., Tiziani D., Funk S., Malyshev D., Nakashima K., van Eldik C., 2019, A&A, 632, A72
Nigro C. et al., 2019, A&A, 625, A10
Observatory Cherenkov Telescope Array, Consortium Cherenkov Telescope Array, 2021, Maier G., ed., CTAO Instrument Response Functions - prod5 version v0.1. Zenodo, Geneva
Olmi B., Torres D. F., 2020, MNRAS, 494, 4357
R Core Team, 2017, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/
Remy Q., Tibaldo L., Acero F., Fiori M., Knödlseder J., Olmi B., Sharma P., 2021, Proc. Sci., 37th International Cosmic Ray Conference. Sissa, Trieste, PoS#886
Renaud M., CTA Consortium, 2011, Mem. Soc. Astron. Italiana, 82, 726
Volpi D., Del Zanna L., Amato E., Bucciantini N., 2008, A&A, 485, 337

APPENDIX A: AKAIKE'S INFORMATION CRITERION COMPARED TO THE TEST STATISTIC

The AIC statistic of a given model with k degrees of freedom is defined by

$$\text{AIC} = 2k - 2\ln(L). \tag{A1}$$

Similarly to equation (3), we can then define:

$$\Delta\text{AIC} = \min\left(\text{AIC}_{\text{Gauss}}, \text{AIC}_{\text{Temp. 1},\alpha}, \text{AIC}_{\text{Temp. 2},\alpha},\right.$$
$$\left.\text{AIC}_{\text{Temp.i},\alpha}\right) - \min(\text{AIC}_{\text{Conf.}}) \tag{A2}$$

A ΔAIC large and positive can thus be related (as in the case of ΔTS) to better prospects for unraveling source confusion. When comparing two different models, we obtain, according to equations (2) and (A1):

$$\text{AIC}_{\text{H}_i} - \text{AIC}_{\text{H}_j} = 2\left(k_{\text{H}_i} - k_{\text{H}_j}\right) - \left(\text{TS}_{\text{H}_i} - \text{TS}_{\text{H}_j}\right). \tag{A3}$$

The latter translates into $\Delta\text{AIC} \approx \Delta\text{TS}$ for the simulations we present. This is due to two reasons. First, because the $H_{\text{Conf.}}$ model of maximum TS (i.e. log-likelihood) results in all cases to be the $H_{\text{Conf.}}$ model of minimum AIC, and the same occurs for the rest of the source templates (i.e. for $H_{\text{Temp. 1},\alpha}$, $H_{\text{Temp. 2},\alpha}$, and $H_{\text{Temp.i},\alpha}$ hypotheses). Secondly, it is due to the difference of free parameters between models ($k_{\text{H}_i} - k_{\text{H}_j}$), which is generally negligible compared to the term concerning the TS and, in most cases (in absolute value), either one or zero. $k_{\text{H}_i} - k_{\text{H}_j} = 1$ occurs for the case of an exponentially cut-off power-law spectral model compared to a simple power-law one (the spatial templates do not have free parameters associated). In any case, $k_{\text{H}_i} - k_{\text{H}_j} \leq 4$ and $k_{\text{H}_i} - k_{\text{H}_j} > 1$ only occur if the Gaussian source hypothesis is that of the smallest AIC compared to all templates (models) of isolated sources in the library, which is highly unusual. From this point, the reader may consider that ΔAIC and ΔTS are virtually switchable throughout this paper. Fig. A1
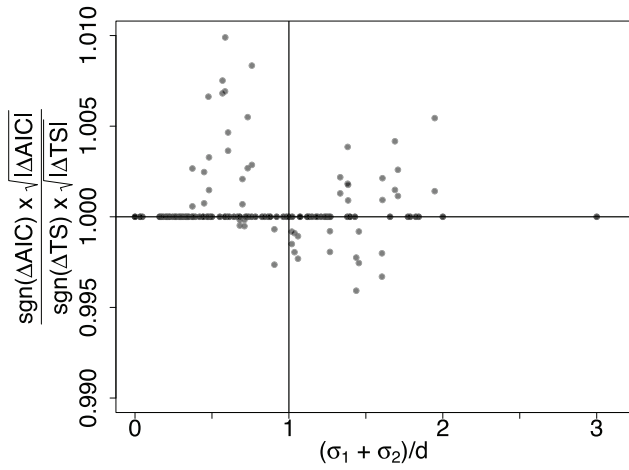
Figure A1. The ratio of ΔAIC and ΔTS defined by equations (A2) and (3) is depicted for different observation simulations.

illustrates how ΔAIC and ΔTS compare for the different simulations performed.

APPENDIX B: TABLES AND PLOTS

This appendix presents different figures and tables to offer a more profound insight into the simulation results and performance. In particular, we illustrate the results for the orientation of the templates (Fig. B1) and the detection significance (Fig. B2), recovered spectral parameters (Figs B3, B4, B5), and flux (Fig. B6) of the sources. Also, we exemplify the performance of the simulations if blurring the source templates (Fig. B8) or modifying the distances to the sources (Fig. B9). Fig. B10 demonstrates the linearity of the simulations (without background). Finally, Table B1 summarizes the computation time required for the simulations.
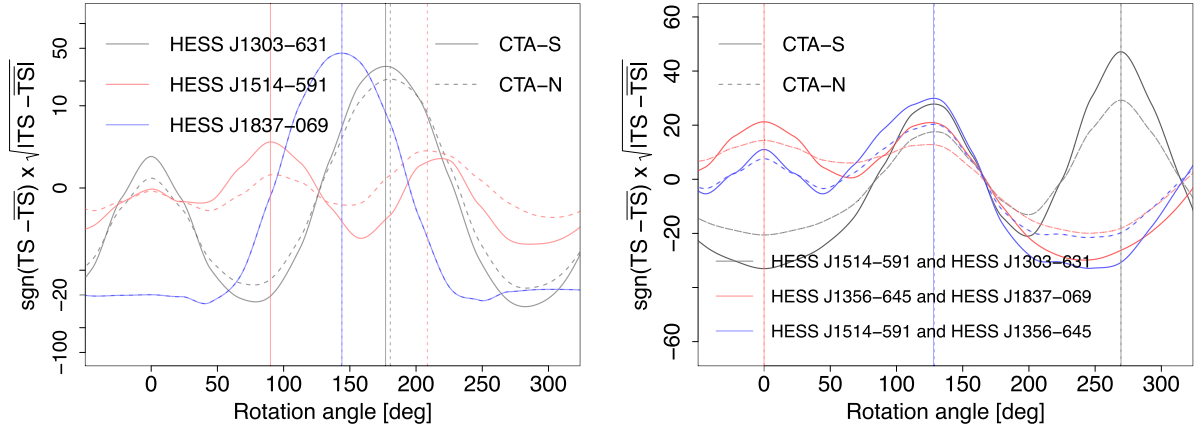
Figure B1. In the left-hand panel, the detection significance retrieved for observation simulations of three sources when isolated (compared to their average over different rotation angles) versus the orientation of the template fitted to the observation simulation. The vertical solid and dashed lines correspond to the absolute maxima for CTA southern and northern arrays, respectively. The input orientations for HESS J1303-631, HESS J1514-591, and HESS J1837-069 in this case are 180°, 216°, and 144°, respectively. In the right-hand panel, the same for different artificially confused sources. The input orientation in the HESS J1514-591 confused with HESS J1303-631 case (in black) is 252°, while in the other two cases (in blue and red) it is 36°.
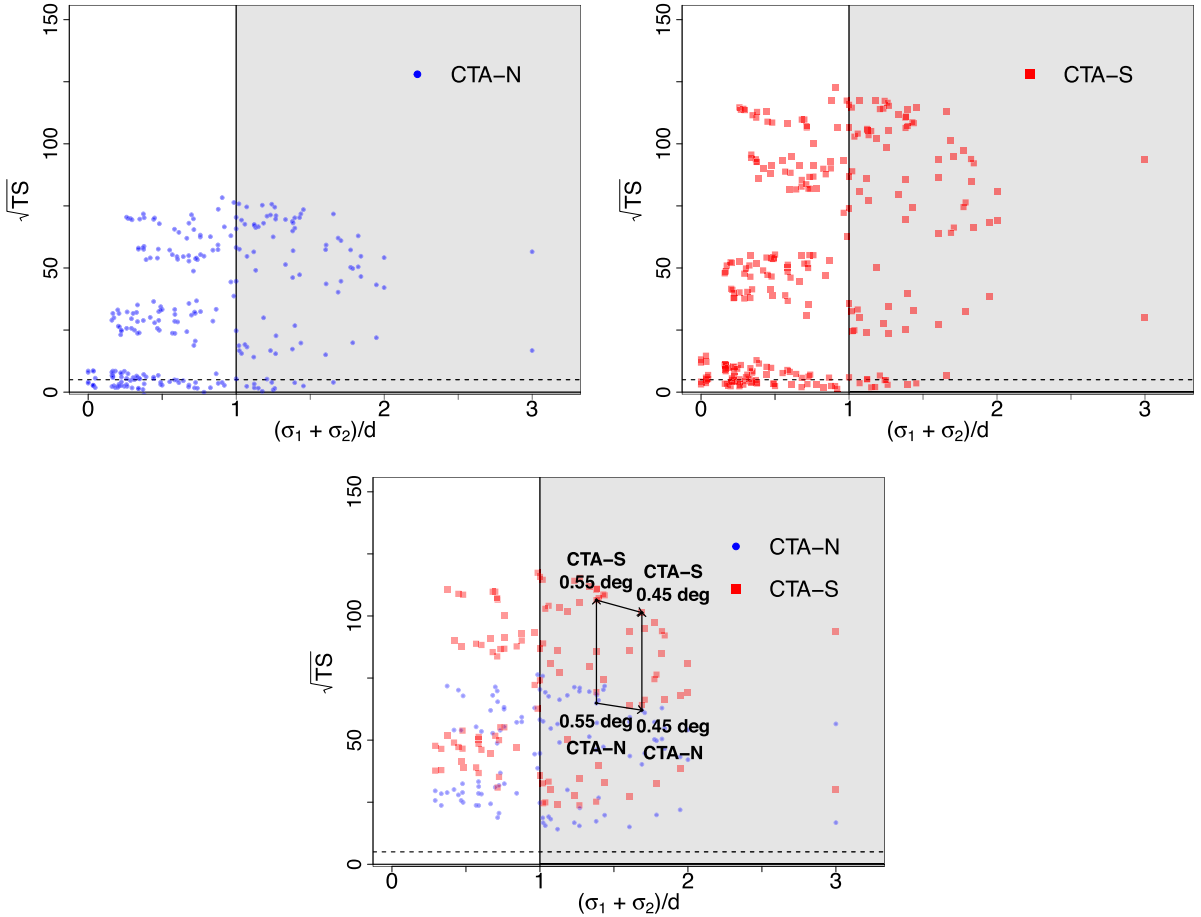
Figure B2. Similarly to Fig. 4, the square root of the TS retrieved for the sources from the best-fitting models to the simulations under the hypothesis of two confused sources, i.e. $TS_{2src,1}$ and $TS_{2src,2}$ in the notation of Section 2.4, compared to the degree of confusion between the sources. We excluded HESS J1554-550 and HESS J1849-000 and the two point-like sources for the lower panel. The dashed horizontal line correspond approximately to a 5σ detection.
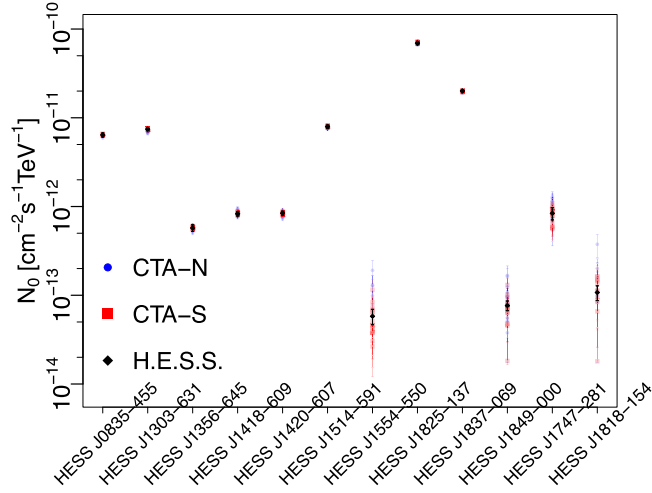
Figure B3. The reconstructed amplitude for the spectrum of each source from the best-fitting models to the simulations under the hypothesis of two confused sources. We exclude the results corresponding to a source detection of TS < 9. The black diamonds represent the H.E.S.S. measurements in Table 1.
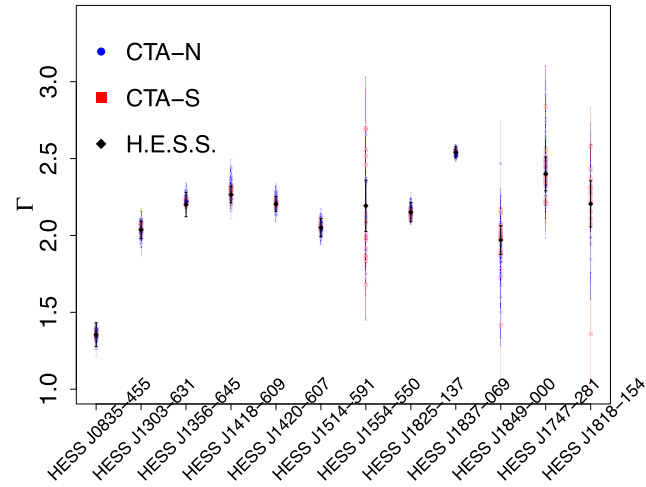

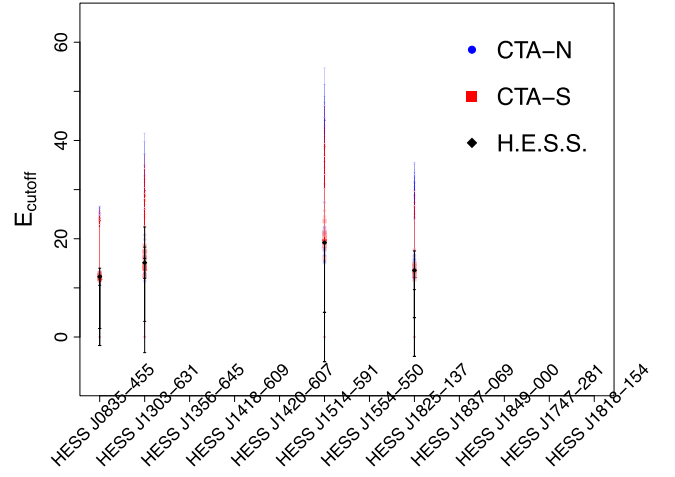**Figure B4.** The same as Fig. B3, but for the spectral index.

Figure B5. The same as Fig. B3, but for the cutoff energy.
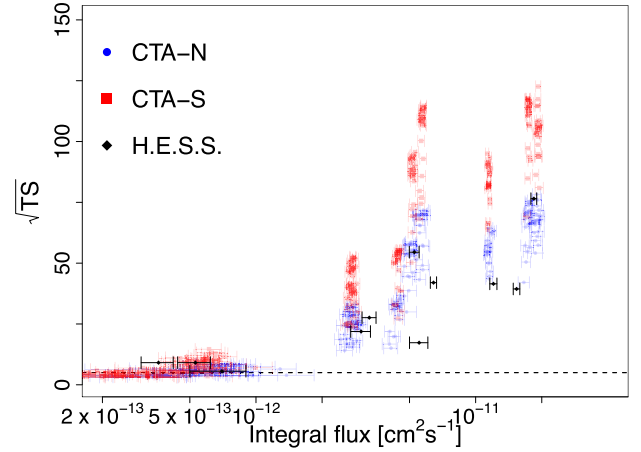

Figure B6. The square root of TS versus the reconstructed integral flux above 1 TeV for each source from the best-fitting models to the simulations under the hypothesis of two confused sources, i.e. $TS_{2src, 1}$ and $TS_{2src, 2}$ in the notation of Section 2.4. The black diamonds represent the H.E.S.S. measurements in Table 1.
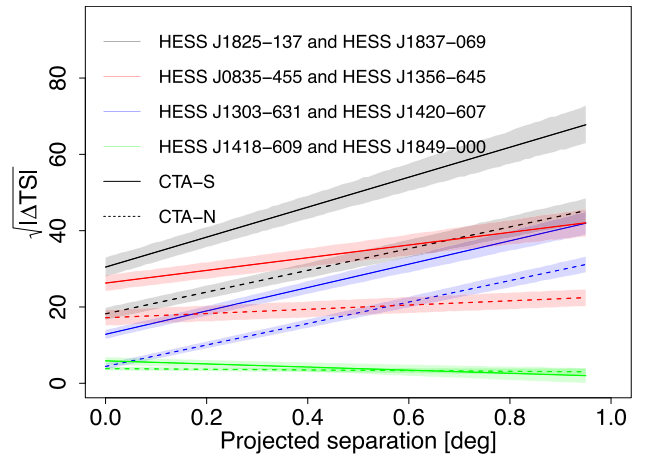

Figure B7. The square root of ΔTS (in absolute value) from equation (3) computed for simulations of several artificially confused PWNe placed at different projected separations from each other.
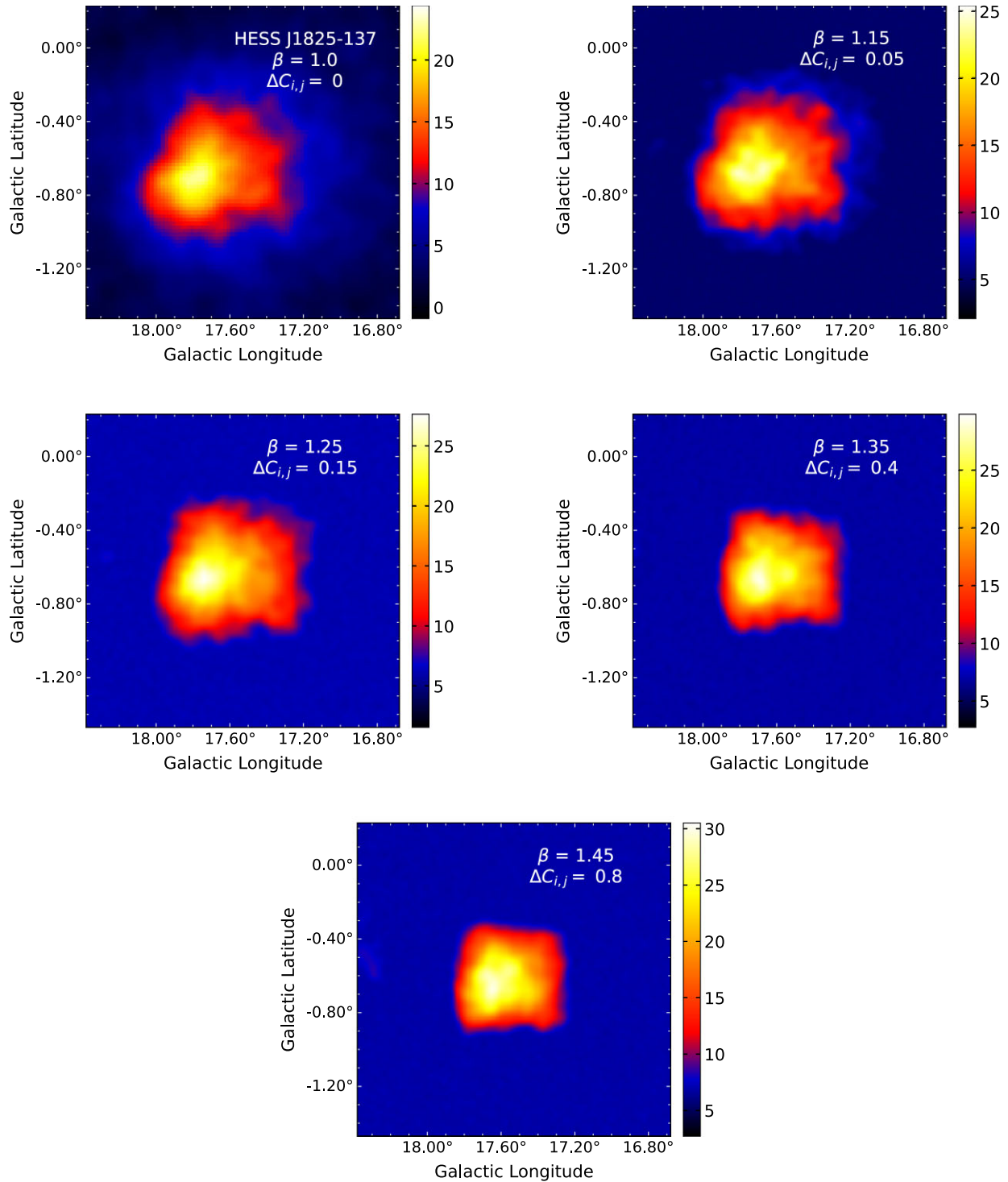
Figure B8. The top left-hand panel depicts an observation simulation of the HESS J1825-137 template blurred with $\beta = 1$, i.e. the softest blurring applied, using 25 h of CTA-S (in events counts without background subtraction). Similarly, the adjacent panels show observation simulations for various deformations of the HESS J1825-137 template corresponding to different values of the β parameter (see Section 3.4), with Pearson's correlation coefficient relative to the unaltered template ($\Delta C_{i,j}$) noted. The plots have been smoothed with a small Gaussian kernel ($\sigma \sim 0.04°$) applied.
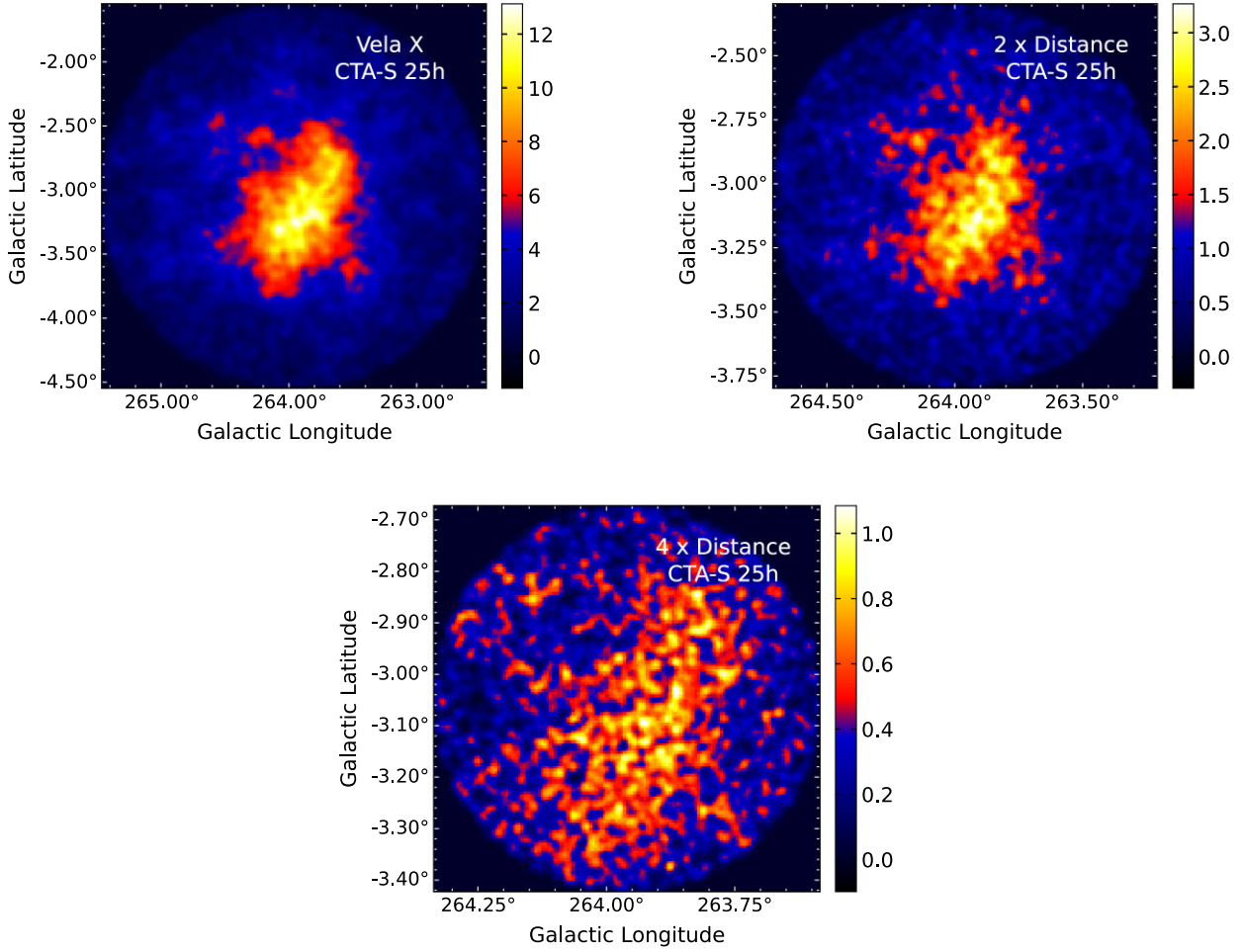
Figure B9. Left-hand panel: sky map resulting from the simulation of the 25-h Vela X observation (see the right-hand panel in Fig. 1) with the CTA southern array above 0.5 TeV. In the central and right-hand panels, we show the same but with the Vela X template rescaled to place it at two and four times its distance. The plots have been smoothed with a small Gaussian kernel ($\sigma \sim 0.04°$) applied.
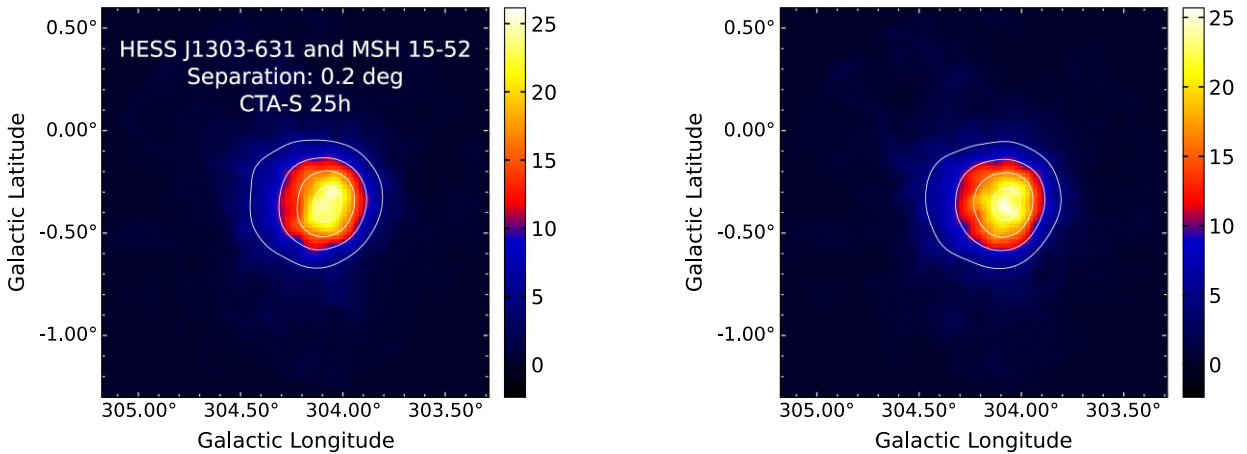
Figure B10. Left-hand panel: a simulation of the HESS J1303-631 and MSH 15-52 templates confused, with $0.2°$ of separation under 25 h of observations with the CTA southern array (with no background simulated). The simulation is centred at the position of HESS J1303-631. On the right, HESS J1303-631 and MSH 15-52 templates were simulated individually (with no background), and next, the simulations were summed. The white contour lines correspond to 5, 10, 15, 20, and 25 events (counts). The plots have been smoothed with a small ($\sigma \sim 0.04°$) Gaussian kernel applied.

Table B1. The computation time (t_{simul}) required for a CTOOLS simulation of two sources (from spatial templates) plus background for different input parameters (using a regular commercial laptop). The latter are the radius of the field-of-view, observation time (t_{obs}), energy threshold (E_{th}), and total flux of the sources above 1 TeV of energy ($F_{>1TeV}$). The bottom part of the table combine increasing observation times and lower energy thresholds, boosting the cost in computation time.

Input parameters				Computation time (s)
Radius (FoV)	t_{obs}	E_{th}	$F_{>1TeV}$	t_{simul}
(deg)	(h)	(TeV)	($10^{-12} \times$ cm^{-2} s^{-1})	(s)
1	1	1	1	2.24 ± 0.01
2	1	1	1	2.24 ± 0.01
10	1	1	1	2.23 ± 0.01
20	1	1	1	2.27 ± 0.04
1	2	1	1	2.25 ± 0.02
1	10	1	1	2.38 ± 0.02
1	100	1	1	3.95 ± 0.05
1	1	0.5	1	2.27 ± 0.01
1	1	0.1	1	2.69 ± 0.02
1	1	0.03	1	3.60 ± 0.07
1	1	1	2	2.26 ± 0.04
1	1	1	10	2.2 ± 0.1
1	1	1	100	2.26 ± 0.05
10	25	0.5	2	3.45 ± 0.03
10	25	0.1	2	14.2 ± 0.5
10	25	0.03	2	34.6 ± 0.9

This paper has been typeset from a TEX/LATEX file prepared by the author.